Purdue University

Purdue e-Pubs

Department of Computer Science Technical Reports

Department of Computer Science

1990

Modified Successive Overrelaxation (MSOR) and Equivalent 2-Step Iterative Methods for Collocation Matrices

Apostolos Hadjidimos

Yiannis G. Saridakis

Report Number:
90-965

Modified Successive Overrelaxation (MSOR)
and Equivalent 2-Step Iterative Methods
for Collocation Matrices

Apostolos Hadjidimos*
Yiannis G. Saridakis**

Computer Sciences Department
Purdue University
Technical Report CSD-TR-965
CAPO Report CER-90-14
March, 1990

ABSTRACT

We consider a class of consistently ordered matrices which arise from the discretization of Boundary Value Problems (BVPs) when the finite element collocation method, with Hermite elements, is used. Through a recently derived equivalence relationship for the asymptotic rates of convergence of the Modified Successive Overrelaxation (MSOR) and a certain 2-step iterative method, we determine the optimum values for the parameters of the MSOR method, as it pertains to collocation matrices. A geometrical algorithm, which utilizes 'capturing ellipse' arguments, has been successfully used. The fast convergence properties of the optimum MSOR method are revealed after its comparison to several well-known iterative schemes. Numerical examples, which include the solution of Poisson's equation, are used to verify our results.

* Department of Mathematics, University of Ioannina, GR-451 10 Ioannina, Greece. Visiting, Department of Computer Sciences, Purdue University, West Lafayette, IN 47907. The work of this author was supported in part by AFOSR grant 88/0243 and by NSF grant CCR-8619817.

** Department of Mathematics and Computer Science, Clarkson University, Potsdam, NY 13676

1. INTRODUCTION

The problem we wish to consider is the iterative solution of certain large and sparse linear systems that are encountered in applications. One such instance, of importance to mathematical software, is the numerical solution of Poisson's equation on a square, with Dirichlet conditions, when the collocation method with Hermite bicubic elements is used.

In recent years, due to the systematic study performed in [13–15] the collocation method has been proven to be a competitive approximation method which is now an integral part of mathematical software for elliptic problems (e.g., ELLPACK [25]). As the resulting, from the discretization, linear system is large and sparse, there is at least one reason (namely storage, cf. [24]) which makes it important to develop iterative methods for collocation matrices.

Relevant results for iterative methods, as it pertains to collocation matrices, may be found in [22, 26, 10]. In particular, in [26, 10], the complete convergence theory for the Extrapolated Jacobi (EJ), Extrapolated Gauss-Seidel (EGS), Successive Overrelaxation (SOR), and Extrapolated SOR (ESOR) (or, equivalently, Extrapolated Accelerated Gauss-Seidel (EAGS)) methods, is included.

Two were the main reasons which motivated us to initiate an investigation for the convergence properties of the MSOR method:

a) For certain choices of the two relaxation parameters of the MSOR, its asymptotic rate of convergence was the same as that of the SOR or EGS,

b) A recently derived [11] equivalence, between the asymptotic rates of convergence of the MSOR and a particular 2-step iterative method, created the opportunity to algorithmically derive the optimum values of their parameters.

In Section 2 we introduce the necessary formalism for the problem which is then used, together with a geometrical algorithm [17, 16] which utilizes the optimum capturing ellipse arguments, in Section 3 for the determination of the optimum values for the parameters of the MSOR and its equivalent 2-step iterative method. In Section 4 we compare the optimum MSOR against the optimum SOR, EGS and EAGS methods. It reveals that the optimum MSOR method is always faster than the optimum SOR and EGS methods, while it competes with the optimum EAGS to win in all cases of practical interest. These results are verified through three example applications which include the numerical solution of the Poisson-Dirichlet problem in the unit square.

2. FORMULATION OF THE PROBLEM

To fix notation, consider the nonsingular linear system

$$Ax = b, \tag{2.1}$$

where $A \in \mathbb{R}^{n,n}$. Writing A as

$$A = D(I - L - U), \tag{2.2}$$

where D is a nonsingular block diagonal matrix and L, U are strictly lower and strictly upper triangular matrices respectively, the associated block Jacobi iteration matrix B is defined by

$$B := I - D^{-1} A = L + U. \tag{2.3}$$

Then, the case of interest is characterized (cf. [22]) by the following set of hypotheses:

H1 the block Jacobi matrix B of (2.3) is consistently ordered *weakly cyclic of index* 2, so that the matrix A of (2.1) is *2-cyclic* (cf. [28]).

H2 Both $\mu = 0$ and $\mu = \pm\, i$ are eigenvalues, of some positive multiplicity, of the block Jacobi matrix B, while $\mu = \pm\, 1$ are not.

H3 All nonzero eigenvalues of B in (2.3) are lying on the circumference of the unit circle.

Based on H1, one may write B of (2.3) as

$$B = L + U := \begin{bmatrix} O_1 & B_1 \\ B_2 & O_2 \end{bmatrix} \tag{2.4}$$

where the matrices O_1 and O_2 are square null matrices of order n_1 and n_2 respectively, with $0 < n_1,\, n_2 < n$. In accordance with the above partitioning of the Jacobi matrix B the MSOR iterative method, as it pertains to the solution of the system (2.1), is described by

$$\left\{ \begin{array}{rcl} x^{(m)} & = & \mathcal{L}_\Omega\, x^{(m-1)} + (I - \Omega\, L)^{-1}\, \Omega c, \qquad m = 0, 1, 2, \cdots \\ \mathcal{L}_\Omega & := & (I - \Omega\, L)^{-1}\, (I - \Omega + \Omega\, U) \\ c & := & D^{-1}\, b \\ \Omega & := & diag(\omega_1\, I_1,\, \omega_2\, I_2) \end{array} \right. \tag{2.5}$$

where I_j denotes the unit matrix of order n_j ($j = 1, 2$), while ω_1, ω_2 ($\neq 0$) are the two relaxation factors of the MSOR method. Clearly, when $\omega_1 = \omega_2 \equiv \omega$ the MSOR reduces to the SOR method with relaxation factor ω.

Let us now consider the double-Jacobi iterative method (cf. [12])

$$x^{(m+1)} = B^2 x^{(m)} + (I + B) D^{-1} b, \quad m = 0, 1, 2, \cdots \tag{2.6}$$

and its associated completely consistent 2-step method

$$\begin{aligned} x^{(m+1)} &= (\hat{\omega}_1 I + \hat{\omega}_2 B^2) x^{(m)} + (1 - \hat{\omega}_1 - \hat{\omega}_2) x^{(m-1)} \\ &+ \hat{\omega}_2 (I + B) D^{-1} b, \quad m = 0, 1, \cdots \end{aligned} \tag{2.7}$$

where $\hat{\omega}_1$, $\hat{\omega}_2$ are real parameters ($\hat{\omega}_2 \neq 0$). With $\hat{\omega}_1$, $\hat{\omega}_2$ satisfying

$$\hat{\omega}_1 = 2 - \omega_1 - \omega_2 \quad \text{and} \quad \hat{\omega}_2 = \omega_1 \omega_2 \ (\neq 0) \tag{2.8}$$

it has recently been shown in [11] that the MSOR method of (2.5) and the 2-step method of (2.7) are *equivalent*, in the sense that their asymptotic rates of convergence are the same. Therefore, the problem of determining the *optimum* values of the parameters ω_1, ω_2 of the MSOR method is equivalent to that of finding the optimum values of the parameters $\hat{\omega}_1$, $\hat{\omega}_2$ of the 2-step method of (2.7) and then determining ω_1 and ω_2 by means of (2.8) or, equivalently, as the roots of the quadratic equation

$$Z^2 - (2 - \hat{\omega}_1) Z + \hat{\omega}_2 = 0. \tag{2.9}$$

Moreover, to comply with known results in the literature, we use the transformation

$$\hat{\omega}_1 = (1 - \omega)(1 + \lambda \psi^2), \quad \hat{\omega}_2 = \omega(1 + \lambda \psi^2) \tag{2.10}$$

and write (2.7) as

$$\begin{aligned} x^{(m+1)} &= (1 + \lambda \psi^2) B_\omega x^{(m)} - \lambda \psi^2 x^{(m-1)} \\ &+ \omega(1 + \lambda \psi^2)(I + B) D^{-1} b, \quad m = 0, 1, \cdots \end{aligned} \tag{2.11}$$

where B_ω is the *extrapolated* double-Jacobi matrix

$$B_\omega := (1 - \omega) I + \omega B^2 = I - \omega (I - B^2). \tag{2.12}$$

At this point we would like to remark that several very interesting results concerning 2-step or in general k-step iterative methods may be found in the literature (e.g., [1–9], [11], [16], [17–21], [23]). The particular method in (2.11) has been analyzed in [17–20], [6], [1], [3–4], [16], [5] and [2]. The treatment in [17–20] and [16] contains the complete analysis for both cases of fixed or varying parameter ω. Following the analysis therein, it is known that the parameters λ and ψ can be seen as functions of the real and imaginary semiaxes, M_R and M_I respectively, of the *capturing ellipse for the spectrum* $\sigma(B_\omega)$ *of* B_ω (that is, an ellipse which is symmetric about both axes and contains $\sigma(B_\omega)$ in its interior). In particular

$$\lambda = \frac{M_R - M_I}{M_R + M_I} \tag{2.13}$$

while ψ is a solution to

$$\lambda \psi^2 - \frac{2}{M_R + M_I}\, \psi + 1 = 0. \tag{2.14}$$

Furthermore, the asymptotic rate of convergence $R_\infty^{(2)}$ of the 2-step method in (2.11) satisfies

$$R_\infty^{(2)} = - \ln \psi. \tag{2.15}$$

Combining (2.13) and (2.14), it is evident that

$$\psi = \frac{M_R + M_I}{1 + \sqrt{1 - M_R^2 + M_I^2}}, \tag{2.16}$$

with *convergence* condition

$$0 < \psi < 1. \tag{2.17}$$

In many of the previous references it is shown that the condition in (2.17) holds if and only if $M_R < 1$ or, equivalently, if and only if the spectrum $\sigma(B_\omega)$ of B_ω lies in the strip S

$$S := \{ z \in \mathbb{C} /\ |Re(z)| < 1 \}, \tag{2.18}$$

that is,

$$\sigma(B_\omega) \subset S. \tag{2.19}$$

By virtue of H1–H3 the eigenvalues of B^2 are lying either on the circumference of the unit circle (but not at 1) or at the origin. Thus, the transformation of (2.12) implies that, for ω in the interval (0,1), the eigenvalues of B_ω are lying in the strip S of (2.18), whence we obtain that the 2-step method of (2.11) *converges* if and only if

$$0 < \omega < 1. \tag{2.20}$$

The problem now of determining the *optimum* asymptotic rate of convergence $R_\infty^{(2)}$ of (2.15) is equivalent to the problem of finding the *optimum* capturing ellipse of B_ω over all ω in (0,1), or, in mathematical terms, equivalent to the solution of the min-max problem

$$\psi = \min_{\omega,\, f_2} \left\{ \max_{\tau} \left\{ \frac{M_R + M_I}{1 + \sqrt{1 - M_R^2 + M_I^2}} \right\} \right\} \tag{2.21}$$

where $f_2 \equiv f^2 := M_R^2 - M_I^2$ and $\tau \in \sigma(B_\omega)$. Observe now that if τ and ξ are eigenvalues of B_ω and $I - B^2$ then, by (2.12),

$$\tau = 1 - \omega\,\xi \equiv 1 - \omega\,(\gamma + i\,\delta), \qquad i^2 = -1. \tag{2.22}$$

The equation of the capturing ellipse for $\sigma(B_\omega)$ which intersects the spectrum $\sigma(B_\omega)$ at the point $\tau = 1 - \omega\,\gamma - i\,\omega\,\delta$ satisfies

$$\frac{(1 - \omega\,\delta)^2}{M_R^2} + \frac{\omega^2\,\delta^2}{M_I^2} = 1$$

or, equivalently,

$$\frac{(d - \gamma)^2}{a^2} + \frac{\delta^2}{b^2} = 1 \tag{2.23}$$

where $\xi = \gamma + i\,\delta$ and

$$d := \frac{1}{\omega}, \quad a := \frac{M_R}{\omega}, \quad b := \frac{M_I}{\omega}. \tag{2.24}$$

Apparently, (2.23) describes an ellipse with center at the point (d,0), real and imaginary semiaxes a and b respectively, which intersects the spectrum of $I - B^2$ at the point $\xi = \gamma + i\,\delta$. Based on these observations the min-max problem in (2.21) can be equivalently written as (cf. [17] or [16])

$$\psi = \min_{\omega,\, c_2} \left\{ \max_{\xi} \left\{ \left| \frac{a + \sqrt{a^2 - c_2}}{d + \sqrt{d^2 - c_2}} \right| \right\} \right\}, \tag{2.25}$$

where $c_2 \equiv c^2 := a^2 - b^2$, and it is equivalent to the problem of finding the optimum "capturing" ellipse for the spectrum of the matrix $I - B^2$. The section that follows is devoted to the solution of this problem.

3. OPTIMUM VALUES

With $\sigma(B)$ denoting the spectrum of the block Jacobi matrix B of (2.3), recalling hypotheses H1–H3, we have that

$$\begin{gathered} \sigma(B) = \left\{ \mu_0^{(m_1)}, \pm\mu_I^{(m_2)}, \pm\mu_j, \pm\frac{1}{\mu_j} \mid j = 1, \ldots, \ell \right\} \\ \ell = \frac{(n - m_1 - 2m_2)}{4} \geq 0 \end{gathered} \tag{3.1}$$

where $\mu_0 := 0$ is an eigenvalue of multiplicity m_1 $(0 < m_1 < n)$, $\mu_I := i$ $(i^2 = -1)$ is an eigenvalue of multiplicity m_2 $(0 < m_2 < n)$ and $\mu_j := \alpha_j + i\,\beta_j$, $(i^2 = -1)$ with

$$\begin{cases} \alpha_j, \beta_j \in I\!R, \quad \alpha_j > 0, \quad \beta_j > 0 \\ \alpha_j^2 + \beta_j^2 = 1 \end{cases} \tag{3.2}$$

for all $j = 1, \ldots, \ell$ when $\ell \neq 0$. Of course when $\ell = 0$ then μ_0 and $\pm\mu_I$ are the only eigenvalues of the matrix B, with multiplicity m_1 and m_2 respectively. Therefore if the matrix $\hat{B}$ is defined by

$$\hat{B} := I - B^2, \tag{3.3}$$

then its spectrum $\sigma(\hat{B})$ is defined by

$$\sigma(\hat{B}) = \left\{ \xi_0^{(m_1)}, \xi_I^{(2m_2)}, \xi_j^{(2)}, \overline{\xi}_j^{(2)} \mid j = 1, \ldots, \ell \right\} \tag{3.4}$$

where ℓ is as in (3.1) and

$$\begin{cases} \xi_0 = 1, \quad \xi_I = 2 \\ \xi_j = (1 - \alpha_j^2 + \beta_j^2) + i(2\alpha_j\,\beta_j) = 2\beta_j^2 + i\,2\alpha_j\,b_j \quad (i^2 = -1) \end{cases} \tag{3.5}$$

and $\overline{\xi_j}$ denotes the complex conjugate of ξ_j. Apparently the eigenvalues of the matrix $\hat{B}$ in (3.3) are lying at the center and on the circumference of the circle $\mathcal{C}$ which is centered at the point (1,0) and has radius 1. Moreover, assuming a counterclockwise ordering of the eigenvalues $\mu_j \mid j = 1, \ldots, \ell$ of the Jacobi matrix on the circumference of the unit circle in the first quadrant, that is

$$1 > \alpha_1 > \alpha_2 > \cdots > \alpha_\ell > 0 \equiv \mathrm{Re}(\mu_I) \tag{3.6}$$

it is evident that the eigenvalues $\xi_j \mid j = 1, \ldots, \ell$ of $\hat{B}$ are ordered in a clockwise fashion on the circumference of the upper half of the circle $\mathcal{C}$, that is

$$0 < 2\beta_1^2 < 2\beta_2^2 < \cdots < 2\beta_\ell^2 < 2 \equiv \mathrm{Re}(\xi_I). \tag{3.7}$$

With this ordering, let the points $P_j(2\beta_j^2, 2\alpha_j\,\beta_j) \mid j = 1, \ldots, \ell$ be the images, in the complex plane, of the eigenvalues $\xi_j \mid j = 1, \ldots, \ell$ of $\hat{B}$. For convenience, let also the points $P_0(1,0)$ and $P_{\ell+1}(2,0)$ be the images in the complex plane of the eigenvalues ξ_0 and ξ_I of $\hat{B}$ respectively, and let the points $P_{2\ell+2-j}(2\beta_j, -2\alpha_j\,\beta_j) \mid j = 1, \ldots, \ell$ be the images of the eigenvalues $\overline{\xi_j} \mid j = 1, \ldots, \ell$ of the matrix $\hat{B}$. Let now H be the polygon with vertices the points $P_j \mid j = k, k+1, \ldots, 2\ell+1$ where $k = 1$ when $2\beta_1^2 \leq 1$ and $k = 0$ when $2\beta_1^2 > 1$. Evidently the polygon H, which is illustrated in Figure 1, is the smallest convex polygon containing the whole spectrum of the matrix $\hat{B}$ in the closure of its interior and is symmetric about the real axis. Of course, when $\ell = 0$, that is when ξ_0, ξ_I are the only eigenvalues of $\hat{B}$, the polygon H reduces to the straight line segment $P_0\,P_{\ell+1} \equiv P_0\,P_1$. The problem now of determining the *optimum* capturing ellipse for the spectrum of the matrix $\hat{B}$ in (3.3) is equivalent to the problem of determining the *optimum* capturing ellipse $\mathcal{E}_H$ for the polygon H. Recall, from Section 2, that the ellipse $\mathcal{E}_H$ is symmetric about the real axis, with equation

$$\frac{[\,\mathrm{Re}(z) - d\,]^2}{a^2} + \frac{[\,\mathrm{Im}(z)\,]^2}{b^2} = 1, \qquad c_2 := c^2 = a^2 - b^2 \tag{3.8}$$

where d, a and b satisfy (2.24) and they are such that

$$\hat{\psi} := \frac{a + \sqrt{a^2 - c_2}}{d + \sqrt{d^2 - c_2}} \tag{3.9}$$

is the solution to the min-max problem of (2.25), that is

$$\hat{R}_\infty^{(2)} = -\ell n\,\hat{\psi} \tag{3.10}$$

is the optimum asymptotic rate of convergence of the 2-step iterative method in (2.11).

We also point out that, from (2.20) and (2.24),

$$d > 1 \tag{3.11}$$

while, from (2.17) and (3.9),

$$a < d. \tag{3.12}$$

Moreover, since the polygon H is symmetric about the real axis, for the determination of the ellipse $\mathcal{E}_H$ it is sufficient to consider the vertices of H with non-negative imaginary part, that is, the vertices $P_j \mid j = k, \ldots, \ell+1$ ($k = 1$ or $k = 0$ when $2\beta_1^2 \le 1$ or $2\beta_1^2 > 1$ respectively). We denote with H^+ the part of H defined by the vertices $P_j \mid j = k, \ldots, \ell+1$ and the positive real semiaxis.

With the notation above, we proceed to determine the optimum capturing ellipse $\mathcal{E}_H$ of the polygon H, by following the algorithm in [16], which clarifies in some sense the algorithm in [17]:

STEP 1 (One-point optimum capturing ellipse.)

Since H^+ can not be a line segment parallel to the imaginary axis, there are no optimum capturing one-point ellipses.

STEP 2 (Two-point optimum capturing ellipse.)

Let

$$\mathcal{E}_{ij},\ i = k,\ k+1, \ldots, \ell \mid j = i+1 \ldots, \ell+1 \mid \qquad k = \begin{cases} 0, & \text{if } 2\beta_1^2 > 1 \\ 1, & \text{if } 2\beta_1^2 \le 1 \end{cases} \tag{3.13}$$

denote the optimum ellipse which intersects H^+ at the points P_i and P_j. We need to determine, if there exist, indices ν_1 and ν_2 such that the ellipse $\mathcal{E}_{\nu_1,\nu_2}$ contains, in the closure of its interior, the positive hull H^+. In such a case $\mathcal{E}_H \equiv \mathcal{E}_{\nu_1,\nu_2}$. For this we consider the following cases:

(i) $\ell = 0$. In this case the spectrum $\sigma(\hat{B})$ of (3.4) consists only of the eigenvalues $\xi_0 = 1$ and $\xi_l = 2$. Hence the positive hull H^+ reduces to the line segment $P_0 P_{\ell+1} \equiv P_0 P_1$. Therefore, the optimum 'ellipse' $\mathcal{E}_{0,1}$ defined by (cf. [17], [16])

$$d := \frac{3}{2}, \quad a := \frac{1}{2}, \quad b := 0, \quad c_2 = a^2 - b^2 = \frac{1}{4}, \tag{3.14}$$

obviously captures H^+ in the closure of its interior whence

$$\mathcal{E}_H = \mathcal{E}_{0,1}. \tag{3.15}$$

(ii) $\ell \neq 0$, **k** $= 1$; $0 < 2\beta_1^2 \leq 1 \longleftrightarrow \frac{1}{2} \leq \alpha_1^2 < 1$. (Figure 1a.) In this case, let us first consider any optimum ellipse $\mathcal{E}_{i,j}$ with $i \neq 1$ (resp. $j \neq \ell + 1$). From this family of ellipses the ones centered at the point $(d_{i,j}, 0)$, with $d_{i,j} \leq 1$, can immediately be discarded from consideration in view of the convergence condition (3.11). For the rest of them we point out that they can intersect the positive hull H^+ only at the points P_i and P_j, since all vertices of H^+ lie on the circumference of the circle $\mathcal{C}$ (see Figure 1a). Consequently, the point P_1 (resp. $P_{\ell+1}$) will either lie strictly in the exterior of all the ellipses $\mathcal{E}_{i,j}$ (when $b_{i,j} > a_{i,j}$; $a_{i,j}$ and $b_{i,j}$ are the real and imaginary semiaxis of $\mathcal{E}_{i,j}$ respectively), or, if P_1 (resp. $P_{\ell+1}$) lies strictly in the interior of some ellipse $\mathcal{E}_{i,j}$ (when $b_{i,j} < a_{i,j}$) then $\mathcal{E}_{i,j}$ will also contain the point (0,0) strictly in its interior, violating in this way the necessary condition for convergence $a_{i,j} < d_{i,j}$ in (3.12). It is therefore evident that none of the ellipses $\mathcal{E}_{i,j}$, with $i \neq 1$ or $j \neq \ell + 1$, is the optimum capturing one. It remains to consider the optimum ellipse $\mathcal{E}_{1,\ell+1}$ which intersects the positive hull H^+ at the points P_1 and $P_{\ell+1}$. Its optimum values are given by (cf. [16]):

$$\begin{cases} d_{1,\ell+1} & := & M_1 + z_0 \\ a_{1,\ell+1} & := & \Big[(z_0 - R_1)(z_0 - R_2) \Big]^{1/2} \\ c_{1,\ell+1}^2 & := & a_{1,\ell+1}^2 \left[1 - \dfrac{R_3}{z_0} \right] \end{cases} \tag{3.16}$$

where z_0 is the unique real root, in the interval $(-\alpha_1^2, 0)$, of the polynomial

$$Q_5(z) = p_1 z^5 + p_2 z^4 + p_3 z^3 + p_4 z^2 + p_5 z + p_6 \tag{3.17}$$

with

$$\begin{cases} p_1 & := & R_1(R_1 + 2R_2 + R_3 - 4R_4) + R_2(R_2 + R_3 - 4R_4) + R_4(4R_4 - 2R_3) \\ p_2 & := & R_1[\,4R_4(R_3 - R_4) + R_2(12R_4 - 5R_3 - 4R_2) + R_1(2R_4 - R_3 - 4R_2)\,] \\ & & + R_2[\,4R_4(R_3 - R_4) + R_2(2R_4 - R_3)\,] - R_3 R_4^2 \\ p_3 & := & R_1\{\,R_2[R_4(2R_4 - 4R_3) + R_2(3R_3 - 4R_4)\,] \\ & & + R_1[R_4^2 + R_2(4R_2 + 3R_3 - 4R_4) - R_1 R_3\,]\,\} + R_2^2\,(R_4^2 - R_2 R_3) \\ p_4 & := & R_1 R_2 R_3\,[\,R_1(3R_1 - 4R_4) + R_2(3R_2 - 4R_4) + 2R_4^2\,] \\ p_5 & := & 3R_1^2 R_2^2 R_3(2R_4 - R_2 - R_2), \\ p_6 & := & R_1^2 R_2^2 R_3(R_1 R_2 - R_4^2) \end{cases} \tag{3.18}$$

and

$$\begin{cases} R_1 := -\dfrac{M_2 M_3}{M_4}, & R_2 := -\dfrac{M_2 M_4}{M_3} \\ R_3 := \dfrac{M_3 M_4}{M_2} & R_4 := -M_1 \end{cases} \tag{3.19}$$

with

$$\begin{cases} M_1 = \dfrac{Re(P_1) + Re(P_{\ell+1})}{2} = \dfrac{2\beta_1^2 + 2}{2} = 1 + \beta_1^2 = 2 - \alpha_1^2 \\ M_2 = \dfrac{Re(P_{\ell+1}) - Re(P_1)}{2} = 1 - \beta_1^2 = \alpha_1^2 \\ M_3 = \dfrac{Im(P_1) + Im(P_{\ell+1})}{2} = \alpha_1 \beta_1 \\ M_4 = \dfrac{Im(P_{\ell+1}) - Im(P_1)}{2} = -\alpha_1 \beta_1 . \end{cases} \tag{3.20}$$

Taking now into consideration the relationships (3.18) – (3.20) we write the polynomial $Q_5(z)$ as

$$Q_5(z) = 4(\alpha_1^2 + 3)(z - \alpha_1^2)^2 \, Q_3(z) \tag{3.21}$$

where

$$Q_3(z) := z^3 + pz^2 + qz + r \tag{3.22}$$

with

$$\begin{cases} p := (1 - \alpha_1^4) / (\alpha_1^2 + 3) \\ q := \alpha_1^2 \, [2 - \alpha_1^2 (1 + \alpha_1^2)] / (\alpha_1^2 + 3) \\ r := \alpha_1^4 (1 - \alpha_1^2)^2 / (\alpha_1^2 + 3) \end{cases} \tag{3.23}$$

By showing that the 'discriminant' Δ of the cubic $Q_3(z)$ in (3.22), defined by

$$\Delta := \frac{\delta_1^2}{4} + \frac{\delta_2^3}{27} \tag{3.24}$$

where

$$\delta_1 := \frac{1}{27}(2p^3 - 9pq + 27r), \quad \delta_2 := \frac{1}{3}(3q - p^2), \tag{3.25}$$

satisfies

$$\Delta > 0 \text{ for } \frac{1}{2} \le \alpha_1^2 < 1 \text{ or } 0 < 2\beta_1^2 \le 1, \tag{3.26}$$

it is evident that z_0 is the unique real root of the cubic $Q_3(z)$. Therefore, by making use of Cardan's formulas, we obtain that

$$z_0 = \left[-\frac{\delta_1}{2} + \Delta^{1/2} \right]^{1/3} + \left[-\frac{\delta_1}{2} - \Delta^{1/2} \right]^{1/3} - \frac{p}{3} \tag{3.27}$$

where δ_1, Δ and p are as in (3.25), (3.24) and (3.23) respectively.

To prove now that the optimum ellipse $\mathcal{E}_{1,\ell+1}$ is the optimum capturing ellipse $\mathcal{E}_H$ it is sufficient to prove that $\mathcal{E}_{1,\ell+1}$ contains the spectrum $\sigma(\hat{B})$ of (3.4) or, equivalently, the positive hull H^+ in the closure of its interior. But, since $\mathcal{E}_{1,\ell+1}$ intersects the circle $\mathcal{C}$ only at the ponts P_1 and $P_{\ell+1}$ (2,0), this is true if and only if

$$d_{1,\ell+1} > 1 \quad \text{or} \quad a_{1,\ell+1} < 1 \tag{3.28}$$

or, by (3.16) and (3.20), if and only if

$$2 - \alpha_1^2 + z_0 > 1 \longleftrightarrow z_0 > \alpha_1^2 - 1. \tag{3.29}$$

Observing, however, that $Q_3(\alpha_1^2 - 1) = -2(1-\alpha_1^2)^2 \ / (\alpha_1^2 + 3) < 0$ while $Q_3(0) = \alpha_1^4 \ (1-\alpha_1^2)^2 \ / (\alpha_1^2 + 3) > 0$, the condition in (3.29) obviously holds proving that

$$\mathcal{E}_H = \mathcal{E}_{1,\ell+1}. \tag{3.30}$$

The optimum parameters are given by (3.16) or, in view of (3.18) – (3.20), by:

$$\begin{aligned} d &:= 2 - \alpha_1^2 + z_0, \\ a &:= \alpha_1^2 - z_0, \\ c_2 \equiv c^2 \equiv a^2 - b^2 &:= a^2 \left[1 + \frac{1-\alpha_1^2}{z_0} \right] \end{aligned} \tag{3.31}$$

where z_0 is as defined in (3.27).

(iii) $\ell \neq 0$, **k = 0;** $1 < 2\beta_1^2 < 2 \longleftrightarrow 0 < \alpha_1 < \frac{1}{2}$ **(Figure 1b).** All optimum ellipses $\mathcal{E}_{i,j}$ except the ones with $(i, j) = (0, 1), (0, \ell+1), (1, \ell+1)$ are easily discarded by following similar arguments to those developed in case (ii) above. And since the optimum ellipse $\mathcal{E}_{0,\ell+1}$ reduces to the line segment $P_0 P_{\ell+1}$ (hence it cannot be the optimum capturing one) there are two remaining optimum ellipses to be investigated: The ellipse $\mathcal{E}_{1,\ell+1}$ and the ellipse $\mathcal{E}_{0,1}$. We proceed our analysis by distinguishing the following two subcases:

(iii.a) $1 < 2\beta_1^2 \leq \frac{3}{2} \longleftrightarrow \frac{1}{4} \leq \alpha_1^2 < 1/2$. In this case the optimum ellipse $\mathcal{E}_{0,1}$ cannot be the optimum capturing one. The reason is that, since by (3.40) – (3.41) below, the abscissa $d_{0,1}$ of its center lies in the interval $(\frac{1+2\beta_1^2}{2}, 2\beta_1^2)$ and since $2\beta_1^2 \leq 3/2$, the point $P_{\ell+1}$ will always lie in the exterior of $\mathcal{E}_{0,1}$. The optimum values for the ellipse $\mathcal{E}_{1,\ell+1}$ are defined in (3.16) – (3.23) and, as Δ of (3.24) satisfies $\Delta > 0$ for all $1/4 \leq \alpha_1^2 < 1/2$, the value of the root z_0 is still given by (3.27). Moreover, we have shown, in case (ii) above, that the ellipse $\mathcal{E}_{1,\ell+1}$ contains the vertices $P_j \mid j = 2, \ldots, \ell$ of the positive hull H^+ in the closure of its interior. Therefore, $\mathcal{E}_{1,\ell+1}$ will be the optimum capturing ellipse $\mathcal{E}_H$ as long as the vertex P_0 of H^+ lies in its interior or, equivalently, if and only if

$$d_{1,\ell+1} \leq 3/2. \tag{3.32}$$

Recalling now the relationships (3.16) and (3.20) the condition in (3.32) above can be written as

$$2 - \alpha_1^2 + z_0 \leq \frac{3}{2} \longleftrightarrow z_0 \leq \alpha_1^2 - \frac{1}{2}, \tag{3.33}$$

which, as z_0 is the unique root of $Q_3(z)$ of (3.22) in the interval $(-\alpha_1^2, 0)$ and $Q_3(0) > 0$, is valid if and only if

$$Q_3(\alpha_1^2 - \frac{1}{2}) \geq 0. \tag{3.34}$$

By a straight-forward calculation we found out that

$$Q_3(\alpha_1^2 - \frac{1}{2}) = \frac{1}{8(\alpha_1^2 + 3)}\, Q_2(\alpha_1^2) \tag{3.35}$$

where

$$Q_2(z) = 4z^2 + z - 1 \tag{3.35}$$

with roots

$$z^{\pm} = \frac{-1 \pm \sqrt{17}}{8}. \tag{3.36}$$

Therefore, the condition in (3.34) or (3.32) is valid if and only if

$$\frac{\sqrt{17} - 1}{8} \le \alpha_1^2 < \frac{1}{2} \longleftrightarrow 1 < 2\beta_1^2 \le \frac{9 - \sqrt{17}}{4} \tag{3.37}$$

in which case

$$\mathcal{E}_H \equiv \mathcal{E}_{1,\ell+1} \tag{3.38}$$

with optimum values defined in (3.31) and (3.27).

We point out that for

$$\frac{1}{4} \le \alpha_1^2 < \frac{\sqrt{17} - 1}{8} \longleftrightarrow \frac{9 - \sqrt{17}}{4} < 2\beta_1^2 \le 3/2 \tag{3.39}$$

there is no optimum capturing two-point ellipse.

(iii.b) $\boldsymbol{\frac{3}{2} \le 2\beta_1^2 < 2 \longleftrightarrow 0 < \alpha_1^2 \le \frac{1}{4}}$. In this case the optimum ellipse $\mathcal{E}_{1,\ell+1}$ cannot be the optimum capturing one, since by (3.16) its center $d_{1,\ell+1}$ lies in the interval $(2\beta_1^2,\ 1 + \beta_1^2)$ and $2\beta_1^2 \ge 3/2$, and therefore the point P_0 will always lie in the exterior of $\mathcal{E}_{1,\ell+1}$. The optimum values for the ellipse $\mathcal{E}_{0,1}$ are given by

$$\begin{cases} d_{0,1} &= M_1 + z_0 \\ a_{0,1} &= \Big[(z_0 - R_1)(z_0 - R_2) \Big]^{1/2} \\ c_{0,1}^2 &= a_{0,1}^2 \left[1 - \dfrac{R_3}{z_0} \right] \end{cases} \tag{3.40}$$

where z_0 is the unique real root in the interval $(0,\ \frac{1}{2} - \alpha_1^2)$ of the quintic polynomial $Q_5(z)$ defined by the relationships (3.17) – (3.19) and, instead of (3.20), by

$$
\begin{aligned}
M_1 &:= \beta_1^2 + \frac{1}{2} = \frac{3}{2} - \alpha_1^2, \\
M_2 &:= \beta_1^2 - \frac{1}{2} = \frac{1}{2} - \alpha_1^2, \\
M_3 &:= M_4 := \alpha_1 \beta_1 > 0.
\end{aligned}
\tag{3.41}
$$

After a modest amount of algebra one may write the polynomial $Q_5(z)$ as

$$
Q_5(z) = \frac{1}{2R_1} (z - R_1)^2 (4R_1^2 + 8R_1 - 1) Q_3(z) \tag{3.42}
$$

where the cubic polynomial $Q_3(z)$ is defined by

$$
Q_3(z) := z^3 + pz^2 + qz + r \tag{3.43}
$$

with

$$
\begin{cases}
p = -(4R_1^2 - 1)(2R_1 + 1) / 2(4R_1^2 + 8R_1 - 1) \\
q = -R_1(R_1 + 1)(4R_1^2 - 1) / (4R_1^2 + 8R_1 - 1) \\
r = R_1^2 (2R_1 - 1)^2 (2R_1 + 1) / 2(4R_1^2 + 8R_1 - 1)
\end{cases}
\tag{3.44}
$$

and R_1 as defined by (3.19) and (3.41), that is

$$
-\frac{1}{2} < R_1 := \alpha_1^2 - \frac{1}{2} = \frac{1}{2} - \beta_1^2 \le -\frac{1}{4}. \tag{3.45}
$$

By showing that the 'discriminant' Δ of the cubic $Q_3(z)$ in (3.43), defined by (3.24) – (3.25) and (3.44), satisfies

$$
\Delta > 0 \quad \text{for all} \quad 0 \le \alpha_1^2 \le \frac{1}{4}, \tag{3.46}
$$

the root z_0 of $Q_5(z)$ is the unique real root of $Q_3(z)$ whence

$$
z_0 = \left[-\frac{\delta_1}{2} + \Delta^{1/2} \right]^{1/3} + \left[-\frac{\delta_1}{2} - \Delta^{1/2} \right]^{1/3} - \frac{p}{3} \tag{3.47}
$$

where δ_1 and Δ are as in (3.24) – (3.25) with p, q and r defined of course in (3.44).

Observe now that as the ellipse $\mathcal{E}_{0,1}$ intersects the circle $\mathcal{C}$ at the point P_1, while the number of intersection points in the upper half plane of these two quadratic curves are at most two, it is evident that the optimum ellipse $\mathcal{E}_{0,1}$ will be the optimum capturing ellipse $\mathcal{E}_H$ as long as the point $P_{\ell+1}$ belongs to the closure of its interior or, equivalently, if and only if

$$d_{0,1} \geq 3/2. \tag{3.48}$$

Recalling now (3.40), (3.41) and (3.45), the condition in (3.48) is equivalent to

$$z_0 \geq \alpha_1^2 \equiv R_1 + \frac{1}{2} > 0 \tag{3.49}$$

which, since by (3.43) – (3.45) there holds $Q_3(0) < 0$ for all values of $R_1 = \alpha_1^2 - \frac{1}{2}$ in $(-\frac{1}{2}, -\frac{1}{4})$, is valid if and only if

$$Q_3(\alpha_1^2) \equiv Q_3(R_1 + \frac{1}{2}) \leq 0. \tag{3.50}$$

This, in turn, is equivalent to

$$\hat{Q}_3 := 2(4R_1^2 + 8R_1 - 1)\, Q_3(\, R_1 + \frac{1}{2}\,) \geq 0, \tag{3.51}$$

as $4R_1^2 + 8R_1 - 1 < 0$ for all R_1 in $(-\frac{1}{2}, -\frac{1}{4})$. By showing now that

$$\hat{Q}_3 = R_1(2R_1 + 1)\,(10R_1 + 3) \tag{3.52}$$

and since $R_1 < 0$, it is clear that (3.51), hence also (3.48), holds if and only if

$$\begin{gathered} -\frac{1}{2} < R_1 \leq -\frac{3}{10} \longleftrightarrow \\ 0 < \alpha_1^2 \leq \frac{1}{5} \longleftrightarrow \frac{8}{5} \leq 2\beta_1^2 < 2. \end{gathered} \tag{3.53}$$

In this case

$$\mathcal{E}_H \equiv \mathcal{E}_{0,1} \tag{3.54}$$

with optimum values given by (3.40) or, equivalently, by

$$\begin{aligned} d &:= \frac{3}{2} - \alpha_1^2 + z_0, \\ a &:= \frac{1}{2} - \alpha_1^2 + z_0, \\ c_2 \equiv c^2 = a^2 - b^2 &:= a^2 \left[1 - \frac{2\alpha_1^2\,(1-\alpha_1^2)}{z_0(1-2\alpha_1^2)} \right] \end{aligned} \tag{3.55}$$

where z_0 is as defined in (3.47).

We point out that for

$$\frac{1}{5} < \alpha_1^2 \le \frac{1}{4} \longleftrightarrow \frac{3}{2} \le 2\beta_1^2 < \frac{8}{5} \tag{3.56}$$

there is no two-point optimum capturing ellipse.

STEP 3 (Three-point optimum capturing ellipse).
This step of the algorithm is necessary only when there is no two-point optimum capturing ellipse, that is in view of (3.39) and (3.56), when

$$\frac{9-\sqrt{17}}{4} < 2\beta_1^2 < \frac{8}{5} \longleftrightarrow \frac{1}{5} < \alpha_1^2 < \frac{\sqrt{17}-1}{8}. \tag{3.57}$$

In such a case let $\mathcal{E}_{i,j,k} \mid i = 0, \ldots, \ell-1 \mid j = i+1, \ldots, \ell \mid k = j+1, \ldots, \ell+1$ denote any such ellipse which intersects the positive hull H^+ at the points P_i, P_j and P_k. Observe now that if $i \neq 0$ then the three points P_i, P_j and P_k will all lie on the circumference of the circle $\mathcal{C}$, forcing $\mathcal{E}_{i,j,k} \equiv \mathcal{C}$ and violating in this way the convergence condition in (3.12) as $a_{i,j,k} = d_{i,j,k} = 1$. On the other hand if $i = 0$ and $j \neq 1$ or $k \neq \ell+1$ then, as the points P_j and P_k are the only intersection points of the positive hull H^+ and the ellipse $\mathcal{E}_{0,j,k}$ which are lying on the circumference of the circle $\mathcal{C}$, it is evident that the vertex P_1 or $P_{\ell+1}$ respectively will always lie in the exterior of the optimum ellipse $\mathcal{E}_{0,j,k}$. Therefore consider the optimum ellipse $\mathcal{E}_{0,1,\ell+1}$ and observe that, as P_1 and $P_{\ell+1}$ are the intersection points of $\mathcal{E}_{0,1,\ell+1}$ and the circle $\mathcal{C}$ while P_0 lies strictly in the interior of the circle $\mathcal{C}$, the arc $P_1\,P_{\ell+1}$ of $\mathcal{C}$ lies in the interior of $\mathcal{E}_{0,1,\ell+1}$. Hence, $\mathcal{E}_{0,1,\ell+1}$ is the optimum capturing ellipse $\mathcal{E}_H$, namely

$$\mathcal{E}_H \equiv \mathcal{E}_{0,1,\ell+1}, \tag{3.58}$$

with optimum values defined by (cf. [17,16]):

$$d = \frac{3}{2}, \quad a = \frac{1}{2}, \quad c_2 \equiv c^2 = a^2 - b^2 := \frac{1}{4(1 - 2\beta_1^2)} = \frac{1}{4(2\alpha_1^2 - 1)}. \tag{3.59}$$

At this point the algorithm terminates.

We conclude this section by summarizing the *optimum values: The optimum rate of convergence* $\hat{R}_\infty^{(2)}$ *of the 2-step iterative method in (2.11)* is given in (3.10) as

$$\hat{R}_\infty^{(2)} = -\ln \hat{\psi} \tag{3.60}$$

where $\hat{\psi}$ is defined in (3.9) as a function of the factors d, a and c_2 of the optimum capturing ellipse $\mathcal{E}_H$ which, in turn, are defined in:

$$\begin{cases} (3.14) & \text{when} \quad \alpha^2 = 0 \ (\ell = 0) \\ (3.55) & \text{when} \quad 0 < \alpha^2 \le \frac{1}{5} \\ (3.57) & \text{when} \quad \frac{1}{5} < \alpha^2 < \frac{\sqrt{17} - 1}{8} \\ (3.31) & \text{when} \quad \frac{\sqrt{17} - 1}{8} \le \alpha^2 < 1. \end{cases} \tag{3.61}$$

In the above, $\alpha := \max\limits_{\mu}\{\mathrm{Re}(\mu)\} \ge 0$ with μ being the eigenvalues of the block Jacobi iteration matrix B of (2.3). The *optimum values for the parameters of the MSOR method of (2.5)* may be found by the following relationships (see Section 2): The optimum asymptotic rate of convergence $R_\infty(\mathcal{L}_\Omega)$ is of course the same as that of the 2-step method of (2.11), hence

$$R_\infty(\mathcal{L}_\Omega) = -\ln \hat{\psi} = \hat{R}_\infty^{(2)} \tag{3.62}$$

and therefore the optimum value of the spectral radius $\rho(\mathcal{L}_\Omega)$ of the MSOR iteration matrix $\mathcal{L}_\Omega$ of (2.5) is apparently given by

$$\rho(\mathcal{L}_\Omega) = \hat{\psi}. \tag{3.63}$$

The optimum values for the relaxation parameters ω_1 and ω_2 of the MSOR method are the roots of the quadratic equation in (2.9), namely

$$\omega_{1,2} = \frac{2 - \hat{\omega}_1 \pm \sqrt{(2 - \hat{\omega}_1)^2 - 4\hat{\omega}_2}}{2}, \tag{3.64}$$

where, by (2.10), (2.13), (2.16) and (2.24)

$$\begin{cases} \hat{\omega}_1 & := & 2(d-1)/(d+\sqrt{d^2-c_2}) \\ \hat{\omega}_2 & := & 2/(d+\sqrt{d^2-c_2}) \end{cases} \tag{3.65}$$

with d and c_2 defined in (3.61). By combining (3.64) and (3.65) it is obtained that

$$\omega_{1,2} = \frac{1}{d+\sqrt{d^2-c_2}} \left[1+\sqrt{d^2-c_2} \pm \sqrt{(d-1)^2-c_2} \right]. \tag{3.66}$$

The optimum values in (3.63) and (3.66) are also shown schematically in Figure 2a and 2b respectively.

4. COMPARISONS AND EXAMPLE APPLICATIONS

In this section we compare the rate of convergence of the optimum MSOR method against the rates of convergence of the optimum SOR, Extrapolated Gauss-Seidel (EGS) and Extrapolated Accelerated Gauss-Seidel (EAGS). The comparisons are performed by direct comparisons of the corresponding spectral radii (Figures 3, 4 and 5), and verified for certain example applications (Table 1). The optimum values for the SOR, EGS and EAGS methods, as it pertains to collocation matrices, have been recently derived in [10]. The numerical results, found in Figures 2–5 and Table 1, for these methods can be also found in [22] and [10]. Throughout this section α is defined by

$$\alpha := \max_{\mu} \{\mathrm{Re}(\mu)\}, \quad \mu \in \sigma(B) \tag{4.1}$$

MSOR versus SOR

Observing that, for the choice $\omega_1 = \omega_2$ of the relaxation parameters, the MSOR method reduces to the SOR, it is evident that, as long as the optimum values for the parameters ω_1 and ω_2 are such that $\omega_1 \neq \omega_2$, the optimum rate of convergence of MSOR will be better than that of the optimum SOR. Inspecting Figure 2b, it is clear that, as $\omega_1 \neq \omega_2$ only for $\alpha = 0$, the optimum MSOR converges faster than the optimum SOR for all α in (0,1), while for $\alpha = 0$ the two optimum methods coincide. This is further demonstrated in Figure 3 which illustrates the numerical comparison of the corresponding spectral radii.

MSOR versus EGS

Let us consider the eigenvalue functional equations for the MSOR and EGS methods, as it pertains to 2-cyclic matrices. In particular, if τ, $\hat{\tau}$ and μ denote the eigenvalues of the MSOR, EGS and Jacobi iteration matrices respectively, then it is known that (cf. [27] and [29–30]) τ and μ satisfy

$$(\tau+\omega_1-1)(\tau+\omega_2-1)=\omega_1\,\omega_2\,\tau\mu^2, \tag{4.2}$$

while $\hat{\tau}$ and μ satisfy (e.g., [10])

$$\hat{\tau}=1-\omega+\omega\mu^2 \qquad (\omega \text{ is the extrapolation factor}). \tag{4.3}$$

Upon setting $\tau=\gamma+i\delta$ and $\hat{\tau}=\hat{\gamma}+i\hat{\delta}$ ($i^2=-1$) and using $|\mu^2|=1$ we obtain

$$\begin{cases} \left[(\gamma+\omega_1-1)^2+\delta^2\right]\left[(\gamma+\omega_2-1)^2+\delta^2\right]=\omega_1^2\,\omega_2^2\,(\gamma^2+\delta^2) \\ \left[(\hat{\gamma}+\omega-1)^2+\hat{\delta}^2\right]=\omega^2, \end{cases} \tag{4.4}$$

while if $\mu^2=0$

$$\begin{cases} \gamma=1-\omega_1 \quad \text{or} \quad \gamma=1-\omega_2 \quad \text{and} \quad \delta=0 \\ \hat{\gamma}=1-\omega \quad \text{and} \quad \hat{\delta}=0. \end{cases} \tag{4.5}$$

Evidently, when one of the two parameters of the MSOR (say ω_2) satisfies $\omega_2=1$, the eigenvalues τ are lying on the circumference or at the center $(1-\omega_1, 0)$ of the circle with radius ω_1. On the other had the eigenvalues $\hat{\tau}$ are of course, lying on the circumference or at the center $(1-\omega, 0)$ of the circle with radius ω. Therefore, when $\omega_2=1$ and $\omega_1=\omega$ the MSOR and the EGS methods have the same asymptotic rate of convergence. Thus, whenever the optimum value of ω_1 and ω_2 is different from 1, the optimum MSOR method will converge faster than the optimum EGS. Inspecting now Figure 2b, one may easily verify that, while $\omega_1<1$ for all α, $\omega_2=1$ when α is approximately 0.3. For this value of α the spectral radii of the MSOR and EGS iteration matrices will be the same, while for all other values of α the spectral radius of the MSOR iteration matrix will be less than the spectral radius of the EGS iteration matrix. This is numerically verified in Figure 4.

MSOR versus EAGS

The comparison of the two optimum methods is performed numerically, by direct comparison of the corresponding spectral radii, and the results are shown in Figure 5. Inspecting Figure 5 one may easily verify that there exists a value $\hat{\alpha}$ (approximately equal to 0.25) such that for $\alpha < \hat{\alpha}$, the optimum EAGS converges faster than the optimum MSOR, while, for $\alpha \geq \hat{\alpha}$ the optimum MSOR method dominates.

We proceed to verify our results for three example linear systems which arise from the discretization of BVPs by the finite-element collocation method. The results are summarized in Table 1.

A. One Dimensional BVPs

Consider the 1-D BVP

$$\begin{cases} c_2\, u''(x) + c_1\, u'(x) + c_0\, u(x) = f(x), & x \in I_x \equiv (0,1) \\ u(0) = u(1) = 0 \end{cases} \tag{4.6}$$

Assuming a uniform partitioning of the interval I_x into N subintervals, we seek an approximate solution u_n, in ;the form

$$u_n(x) = \sum_{k=1}^{n} \delta_k\, \phi_k(x), \quad n = 2N, \quad \phi_k(x) \equiv \text{Hermite cubics}.$$

Using the collocation method (at the Gaussian points) for discretization, one arrives at a linear system (for the unknowns δ_k) whose coefficient matrix A, for specific values of c_0, c_1 and c_2 in (4.6), has the form (e.g., [26], [10])

$$A = \begin{bmatrix} b_2 & b_3 & -b_4 & & & & & & \\ b_4 & b_1 & -b_2 & & & & & & \\ & b_1 & b_2 & b_3 & -b_4 & & & & \\ & b_3 & b_4 & b_1 & -b_2 & & & & \\ & & & & \ddots & & & & \\ & & & & & b_1 & b_2 & b_3 & -b_4 \\ & & & & & b_3 & b_4 & b_1 & -b_2 \\ & & & & & & b_1 & b_2 & -b_4 \\ & & & & & & b_3 & b_4 & -b_2 \end{bmatrix} . \tag{4.7}$$

Example 1 (Interpolation Problem) $c_2 = c_1 = 0$, $c_0 = 1$,

$$b_1 = \frac{9+4\sqrt{3}}{18}, \quad b_2 = \frac{3+\sqrt{3}}{36}, \quad b_3 = \frac{9-4\sqrt{3}}{18}, \quad b_4 = \frac{3-\sqrt{3}}{36}.$$

Example 2 $c_2 = 1$, $c_1 = c_0 = 0$,

$$b_1 = -2\sqrt{3}, \quad b_2 = -(\sqrt{3}+1), \quad b_3 = 2\sqrt{3}, \quad b_4 = \sqrt{3}-1.$$

Two Dimensional BVPs

As a model problem in the 2-D case we consider Poisson's equation in the unit, square with Dirichlet boundary conditions. That is,

$$\begin{cases} \Delta^2 u = f, & \text{on } R := (0,1)\times(0,1) \\ u = g, & \text{on } \partial R. \end{cases} \tag{4.8}$$

Assuming a uniform grid with spacing $h := N^{-1}$, where N is as defined in the 1-D case, we seek an approximate solution u_n in the form

$$u_n(x,y) = \sum_{k=1}^{n} \delta_k \, \phi_k(x,y), \quad n = 4N^2, \quad \phi_k(x,y) \equiv \text{Hermite bicubics}.$$

In analogy with the 1-D case, the collocation produces a linear system whose coefficient matrix A has the form (cf. [22])

$$A = \begin{bmatrix} A_2 & A_3 & -A_4 & & & & & & \\ A_4 & A_1 & -A_2 & & & & & & \\ & A_1 & A_2 & A_3 & -A_4 & & & & \\ & A_3 & A_4 & A_1 & -A_2 & & & & \\ & & & & \cdot & & & & \\ & & & & & \cdot & & & \\ & & & & & & \cdot & & \\ & & & & & A_1 & A_2 & A_3 & -A_4 \\ & & & & & A_3 & A_4 & A_1 & -A_2 \\ & & & & & & & A_1 & A_2 & -A_4 \\ & & & & & & & A_3 & A_4 & -A_2 \end{bmatrix} \tag{4.9}$$

where each $A_i \mid i = 1, 2, 3, 4$ is a $2N \times 2N$ matrix in the form given in (4.7). The corresponding values $b_j^{(i)} \mid j = 1, 2, 3, 4$ for each A_i maybe found in [22].

The above examples have been chosen so that we can be able to demonstrate all possible cases discussed earlier on. In Example 1, the value of $\alpha = \max\{\mathrm{Re}(\mu)\}$ remains less than 1/4 so that the optimum MSOR, although it converges faster than the optimum SOR and EGS methods, is slower than the optimum EAGS. In Example 2 the MSOR method dominates. Example 3 represents a model for elliptic BVPs and is of practical interest. Here the value of α is greater than 1/2 for $N \geq 4$, and therefore the optimum MSOR has the fastest asymptotic rate of convergence.

References

1. Avdelas, G., "A Second Order Stationary Scheme for Complex Linear Systems", *Intern. J. Computer Math.* **14**(1983), 171-181.
2. Avdelas, G., J. de Pillis, A. Hadjidimos and M. Neumann, "A Guide to the Acceleration of Iterative Methods Whose Iterative Matrix is Nonnegative and Convergent", *SIAM J. Matrix Anal. Appl.* **9**(1988), 329-342.
3. Avdelas, G., S. Galanis and A. Hadjidimos, "On the Optimization of a Class of Second Order Iterative Schemes", *BIT* **23**(1983), 50-64.
4. Avdelas, G., and A. Hadjidimos, "Optimum Second Order Stationary Extrapolated Iterative Schemes", *Math. Comput. Simulation* **XXV**(1983), 189-198.
5. Avdelas, G., and A. Leontitsis, "A Method for the Improvement of the Convergence Rates of Stationary Iterative Schemes for the Solution of Complex Linear Systems", *J. Comp. Appl. Maths.* **15**(1986), 1-11.
6. de Pillis, J., "How to Embrace Your Spectrum for Faster Iterative Results", *Linear Algebra Appl.* **34**(1980), 125-143.
7. de Pillis, J. and M. Neumann, "Iterative Methods with k-Part Splittings", *IMA J. Numer. Anal.* **1**(1981), 65-79.
8. Galanis, S., A. Hadjidimos and D. Noutsos, "On the Equivalence of the k-Step Iterative Euler Methods and Successive Overrelaxation Methods for k-Cyclic Matrices", *Math. Comput. Simulation* **30**(1988), 213-230.
9. Golub, G.H. and R.S. Varga, "Chebyshev Semi-Iterative Methods, Successive Overrelaxation Methods, and Second-Order Richardson Iterative Methods", *Numer. Math. Parts I and II* **3**(1961), 147-168
10. Hadjidimos, A., T.S. Papatheodorou and Y.G. Saridakis, "Optimal Block Iterative Schemes for Certain Large, Sparse, and Nonsymmetric Linear Systems", *Linear Algebra Appl.* **110**(1988), 285-318.
11. Hadjidimos, A. and A.K. Yeyios, "Some Recent Results on the Modified SOR Theory," *Linear Algebra Appl.*, to appear.
12. Hageman, L.A. and D.M. Young, *Applied Iterative Methods*, Academic Press, New York, 1981.
13. Houstis, E.N., R.E. Lynch, T.S. Papatheodorou and J.R. Rice, "Evaluation of Numerical Methods for Elliptic Partial Differential Equations", *J. Comput. Phys.* **27**(1978), 323-350.

14. Houstis, E.N., W.F. Mitchell and T.S. Papatheodorou, "A C^1-Collocation Method for Mildly Nonlinear Elliptic Equations on General 2-D Domains", *Advances in Computer Methods for Partial Differential Equations* (R. Vichnevetsky and R.S. Stepleman, eds.), IMACS III, pp. 18-27, 1979.

15. Houstis, E.N., W.F. Mitchell and T.S. Papatheodorou, "Performance Evaluation of Algorithms for Mildly Nonlinear Elliptic Problems", *Internat. J. Numer. Meth. Engrg.* **19**(1983), 665-709.

16. Leontitsis, A., *A Stationary Second Order Iterative Method for the Solution of Linear Systems*, (in Greek), Ph.D. Dissertation, Department of Mathematics, University of Ioannina, Ioannina, Greece, 1983.

17. Manteuffel, T.A., "An Iterative Method for Solving Nonsymmetric Linear Systems with Dynamic Estimation of Parameters", UIUCSD-R-758, Department of Computer Science, University of Illinois, Urbana, IL, 1975.

18. Manteuffel, T.A., "The Tchebychev Iteration for Nonsymmetric Linear Systems", *Numer. Math.* **28**(1977), 307-327.

19. Manteuffel, T.A., "Adaptive Procedure for Estimating Parameters for the Nonsymmetric Tchebychev Iteration", *Numer. Math.* **31**(1978), 183-208.

20. Manteuffel, T.A., "Optimal Parameters for Linear Second-Degree Stationary Iterative Methods", *SIAM J. Numer. Anal.* **19**(1982), 833-839.

21. Niethammer, W. and R.S. Varga, "The Analysis of *k*-Step Iterative Methods for Linear Systems from Summability Theory", *Numer. Math.* **41**(1983), 177-206.

22. Papatheodorou, T.S., "Block AOR Iteration for Nonsymmetric Matrices", *Math. Comp.* **41**(1983), 511-525.

23. Parsons, B.N., "General *k*-Part Stationary Iterative Solutions to Linear Systems", *SIAM J. Numer. Anal.* **24**(1987), 188-198.

24. Rice, J.R., *Matrix Computations and Mathematical Software*, McGraw-Hill, New York, 1981.

25. Rice, J.R. and R.F. Boisvert, *Solving Elliptic Problems Using ELLPACK*, Springer-Verlag, New York, 1985.

26. Saridakis, Y.G., *Parallelism, Applicability and Optimality of Modern Iterative Methods*, Ph.D. Dissertation, Department of Mathematics and Computer Science, Clarkson University, Potsdam, NY, 1985.

27. Taylor, P.J., "A Generalization of Systematic Relaxation Methods for Consistently Ordered Matrices", *Numer. Math.* **13**(1969), 377-395.

28. Varga, R.S., *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1962.

29. Young, D.M., "Convergence Properties of the Symmetric and Unsymmetric Successive Overrelaxation Methods and Related Methods", *Math. Comp.* **24**(1970), 793-807.

30. Young, D.M., *Iterative Solution of Large Linear Systems*, Academic Press, New York, 1971.

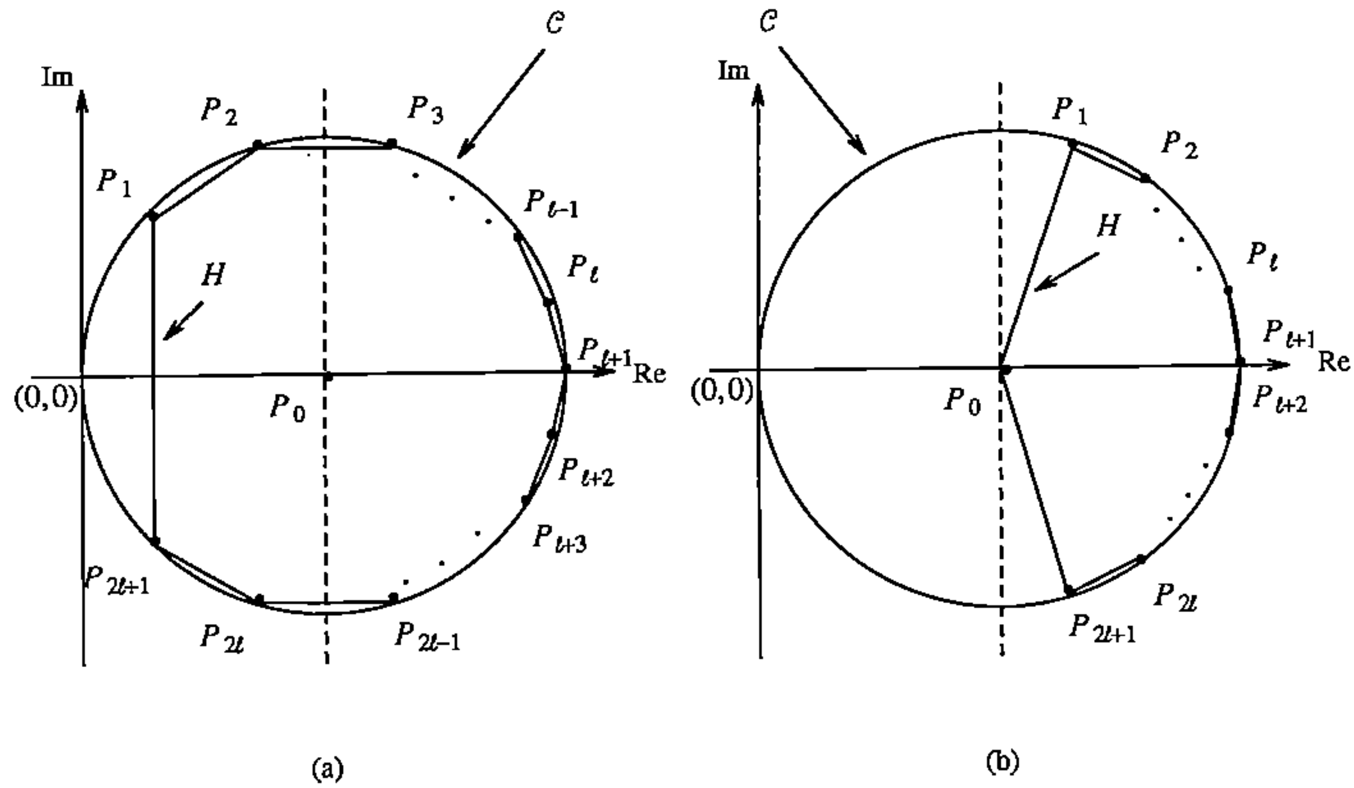

Figure 1. The convex polygon H when (a) $2\beta_1^2 \le 1$, and (b) $2\beta_1^2 > 1$.

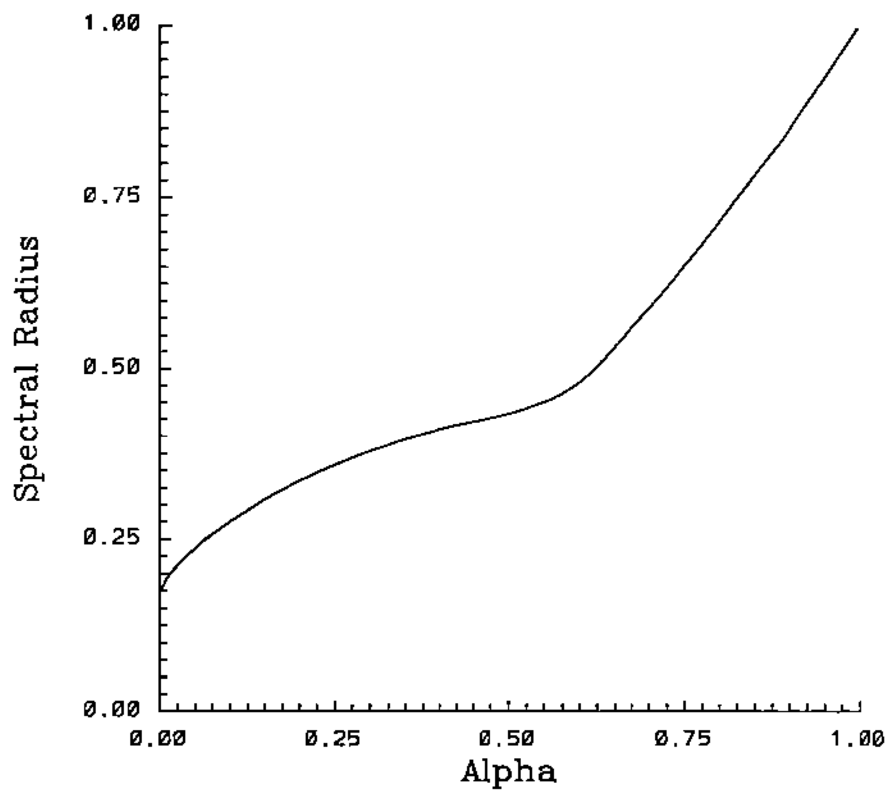

Figure 2a. Optimum values, for the parameters of the MSOR method, as functions of $\alpha := \max\{\mathrm{Re}(\mu)\}$: Spectral Radius $\rho(\mathcal{L}_\Omega)$

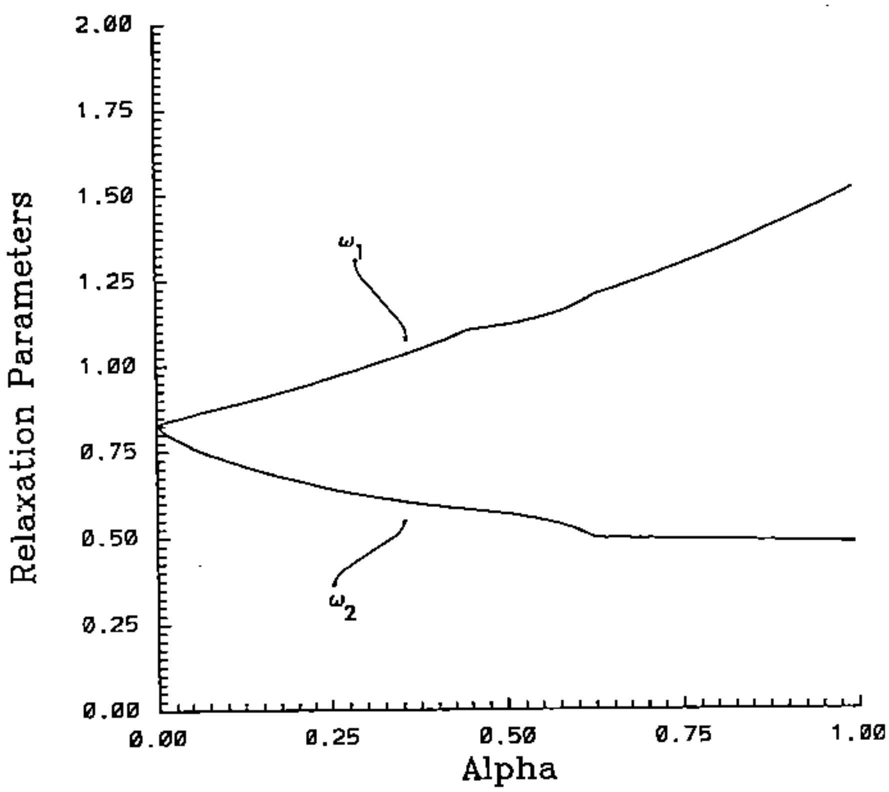

Figure 2b. Optimum values, for the parameters of the MSOR method, as functions of $\alpha := \max\{Re(\mu)\}$: Relaxation parameters ω_1 and ω_2.

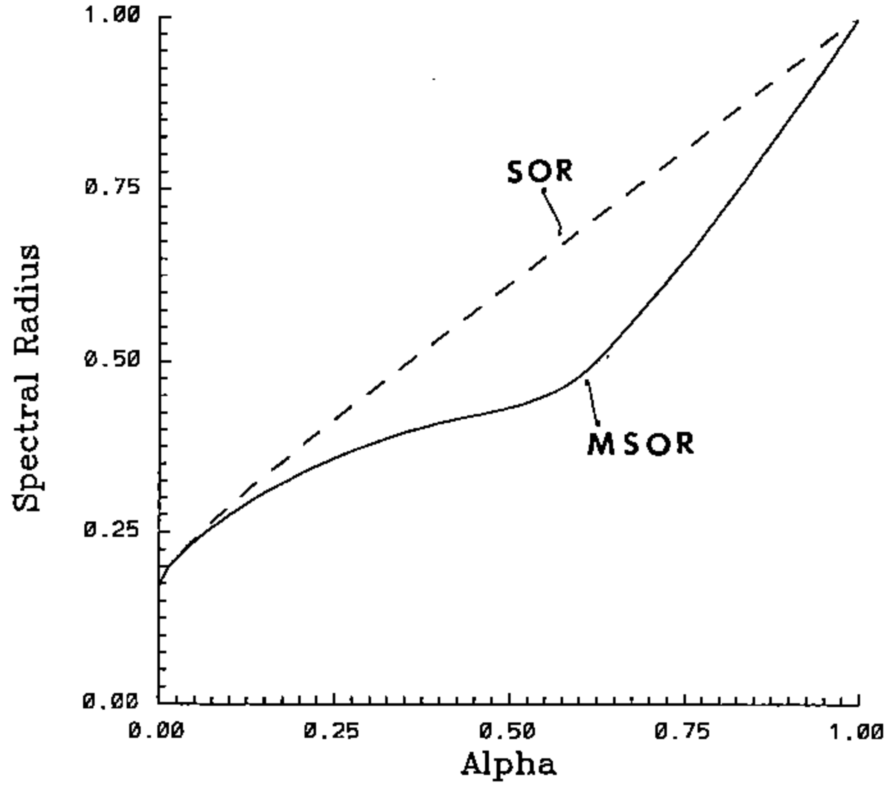

Figure 3. Spectral radii for the optimum MSOR and SOR methods as functions of $\alpha = \max\{\mathrm{Re}(\mu)\}$.

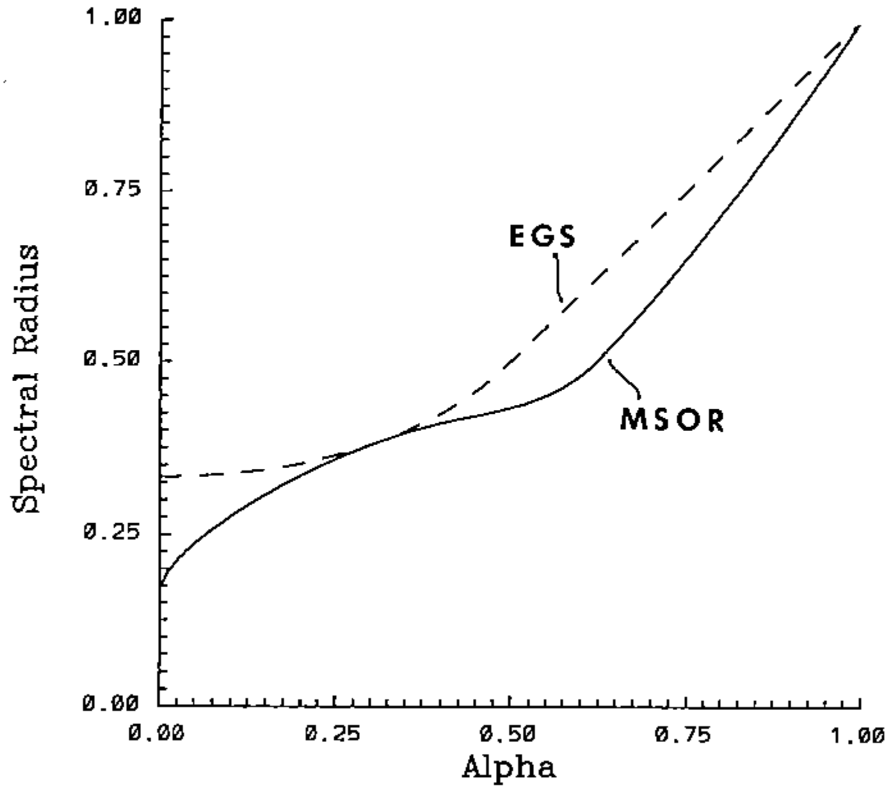

Figure 4: spectral radii for the optimum MSOR and EGS methods as functions of $\alpha = \max\{Re(\mu)\}$.

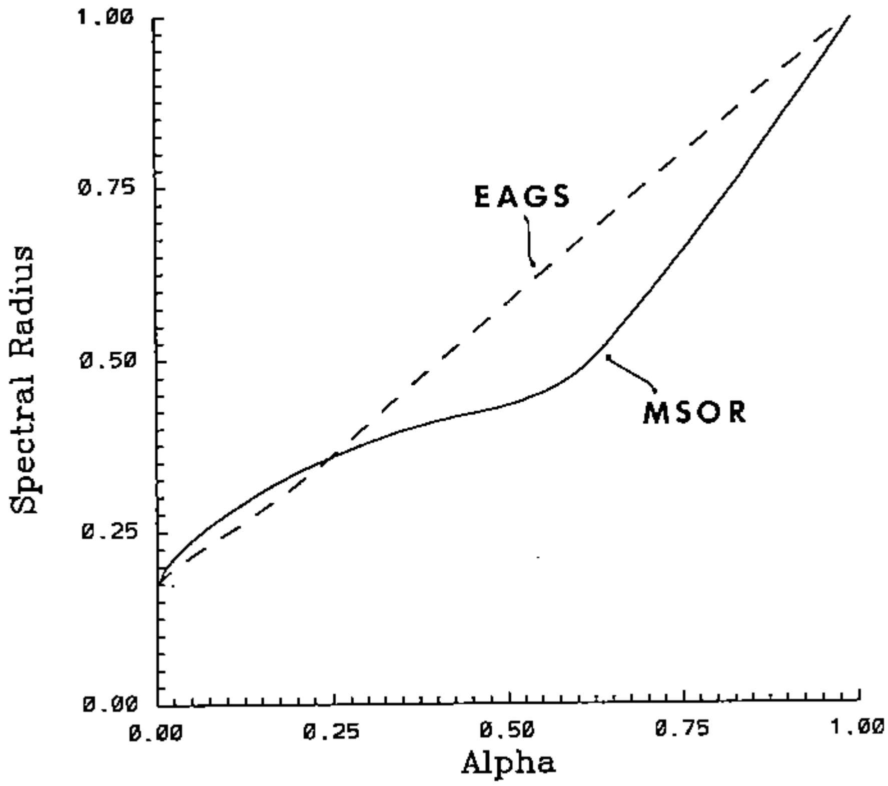

Figure 5. Spectral radii for the optimum MSOR and EAGS methods as functions of $\alpha = \max\{Re(\mu)\}$.

Table 1
Optimum Values from Example Applications

BVP	Number of Subintervals N	$\alpha = \max\{Re(\mu)\}$	SOR		EGS		EAGS			MSOR		
			ω	$\rho(\mathcal{L}_\omega)$	ω	$\rho(\mathcal{L}_{1,\omega})$	r	ω	$\rho(\mathcal{L}_{r,\omega})$	ω_1	ω_2	$\rho(\mathcal{L}_\Omega)$
$u = f$	$N = 4$	0.10102	0.7980	0.2897	0.6621	0.3379	0.7980	0.7517	0.2483	0.8820	0.7237	0.2763
	$N = 8$	0.13198	0.7928	0.3171	0.6587	0.3413	0.7928	0.7312	0.2688	0.8976	0.7033	0.2967
	$N = 16$	0.14011	0.7915	0.3241	0.6577	0.3423	0.7915	0.7256	0.2744	0.9019	0.6983	0.3017
	$N = 32$	0.14217	0.7912	0.3259	0.6574	0.3426	0.7912	0.7742	0.2758	0.9029	0.6970	0.3030
$u'' = f$	$N = 4$	0.70711	0.7441	0.7741	0.5	0.70711	0.7441	0.5780	0.7559	1.2604	0.4946	0.5983
	$N = 8$	0.92388	0.7348	0.9414	0.5	0.92388	0.7348	0.5808	0.9368	1.4428	0.4857	0.8862
$\Delta^2 u = f$	$N = 4$	0.53383	0.7537	0.6397	0.5	0.53383	0.7537	0.6048	0.6092	1.1294	0.5564	0.4436